Journal of Phonetics (1996) **24,** 491–511

Letter to the Editor

The discreteness of phonetic elements and formal linguistics: response to A. Manaster Ramer

Robert F. Port
Department of Linguistics, Department of Computer Science, Program in Cognitive Science, Indiana University, Bloomington, IN 47405, U.S.A.

The phenomenon of ''incomplete neutralization'' and the subtlety of this incompleteness reveal vividly that speech sounds do not fall into discretely distinct phonetic types, and also that auditory impressions cannot be relied upon to identify them. Professor Manaster Ramer appreciates that these failures pose a serious problem for phonology. The reason is that the segmental units of standard phonology inherit the properties of discreteness and auditory identifiability from universal phonetics. So if phonetics is not always discrete, and practical identification is inherently unreliable, then phonology must restructure itself from the ground up. Phonology would then have no way to account for its discrete phonological objects (whether phonemes, rules, syllable trees, metrical grids, constraint rankings, allophonic rules, etc.). This seems to be what the author calls the theoretical bad dream concerning Professor Manaster Ramer. The author argues that all this is true and that an explanation for the discreteness of phonology must be sought elsewhere, not in a hypothetical universal phonetic inventory. In fact, the explanation for phonological discreteness must be sought in the same place it is sought in other sciences (e.g., astronomy, meteorology, economics, physics, biology, etc.)—in the dynamically created structures often exhibited by systems with many degrees of freedom and a constant flow of energy.

1. Introduction

Are phonetic units discretely different from each other? Can careful listeners identify speech sounds reliably? These are the fundamental issues raised by Professor Manaster Ramer (AMR) in his letter to the editor in this issue (Manaster Ramer, 1996). The author is concerned about a phenomenon reported a number of times in the phonetics literature over the past decade. He is apparently alarmed that these results may undermine the methodological and theoretical basis of phonology. In this essay, I will first review the phenomenon itself and agree with AMR that ''incomplete neutralization'' is just a slight variant of many other familiar phenomena from experimental phonetics and phonology, and that it is troubling. Many

0095-4470/96/040491 + 21 $25.00/0

widely attested phenomena offer strong evidence that the set of sounds of human speech, that is, the universal inventory of phonetic categories, cannot be assumed to fall into discretely distinct types that can be reliably differentiated by a first language learner and by a trained linguist or phonetician. Of course, often speech sounds are obviously different (e.g., the initial stops in *bad vs. pad* or the vowels in *bad vs. bed*). But can they *always* be assumed to be discretely different from each other and sufficiently distinct that a careful listener could hear the difference? Traditional linguistic phonetics says yes, but I think the evidence makes this very unlikely. That is, if you ask regarding two short sound bites ''Are these two instances of the same sound or different ones?'', one simply cannot assume that there will always be a correct or consistent answer. But if this is so, then one cannot use phonetics to explain the discreteness of *bad* and *pad* either. The evident discreteness of phonology will need to be accounted for in some other way.

AMR is quite correct to be concerned about the issues of the *Failure of Phonetic Discreteness* and the *Failure of Auditory Identifiability*—about the fact that neither a linguist nor a native speaker can necessarily identify discretely and confidently which phonetic elements they observe in a sample of speech. This failure raises distressing questions about how the discipline of linguistics should obtain data regarding the phonological systems of human language. The phenomenon of incomplete neutralization *is* a worrisome chink in the dam that supports all of current symbol-based phonological theory—just as AMR seems to fear.

Many phonologists prefer to act as though one could assume confidently that the process of careful listening followed by phonetic transcription (at least when performed by a professional) ''makes available'' to the phonologist a representation of *all possible linguistically relevant aspects of a sample of speech.* From this discrete symbolic description, the phonologist constructs various data structures, from phonological segments to syllabic trees to metrical grids. Without this discreteness and the presumed positive identifiability of these units, phonology (as well as the language-learning child) would appear to have no place to begin work. How could phonology account for symbolic structures like language-specific consonant and vowel types, allophones, stress markings, constraint hierarchies, etc.? Where could these complex cognitive objects possibly come from?

Chomsky & Halle (1968) offered a simple answer. They proposed that the universal phonetic space is a discrete alphabet. There is some list of features that is the maximum set of linguistically relevant sound types in languages of the world. All members of our species are born with the ability to reliably and almost effortlessly identify these perceptual objects. These static features serve as units for language-learning children and constrain the possible speech sounds of language such that they take the form of a two-dimensional matrix of feature values *vs.* discrete time. (Although Chomsky & Halle asserted that the features might have continuous values rather than discrete ones, linguists invariably treat them as though they had discrete, quantal values.) Thus, surprisingly, the data structures of phonological descriptions *inherit their discreteness from the phonetic atoms* of which they are constructed. A fundamental discreteness is built into linguistic ontology by Chomsky & Halle. Whatever the physical world might be like, it is assumed that as far as language is concerned, nothing other than discrete, abstract phonetic objects need to be the concern of linguists. This premise seems to be closely related to the Cartesian view that the human mind is quite a different sort of entity from the physical world.

For phoneticians, on the other hand, the failure of phonetic discreteness is in fact rather familiar. Phoneticians have long observed that phonetic objects are discretely different only in ideal circumstances (Lisker & Abramson, 1971; Klatt, 1976; Lindblom, 1983; Keating, 1985). However, although AMR is concerned about the partial neutralization problem and suggests that great amounts of experimental work might have to precede any linguistic analysis, he does not seem to put his finger on the magnitude of the theoretical threat. In this essay, I will try to clarify what is at issue and why it is important. Finally, I will suggest how a theoretical solution to the problem of phonetic discreteness may be found.

2. What is "incomplete neutralization"?

The phenomenon that concerns AMR can be illustrated with the German stems *Bund-* "brotherhood" and *bunt-* "colorful". If the stems have a suffix, as in *Bunde* and *bunte,* then they are pronounced [bʊndə] and [bʊntə]. The orthographic difference between the D and T is clearly appropriate, and illustrates one of the most striking and profound properties of human language: the use of discretely distinct sound types for "spelling" lexical items. These phonological types, the D and T, recur in many different vocabulary items and often may appear in a range of syllable positions. However, when the /d/ or /t/ occurs as the last segment in a syllable (that is, for these stems, when there is no suffix), the contrast is often said to be "neutralized" since both words are pronounced as approximately [bʊnt] with a final [t]. The D/T distinction appears to be lost since both pronunciations (for these words and for many other pairs of voiced and voiceless obstruents in Standard German) merge into a phonetic [t] (or the cognate voiceless stop, fricative or affricate). The traditional linguistic description of this phenomenon is to say that a "neutralization process" has applied that turned the [+voice] feature of the D into the [−voice] feature of the T. In the more recent optimality theory formulation of such data, German would be said to rank certain simultaneous constraints differently than, say, English. But the simultaneous constraints are still stated in terms of standard phonetic features.

The surprising experimental result is that if you do either a production experiment or a perception experiment using a reasonable number of tokens of each minimal pair, you find that the two sets of [t]-like sounds are consistently somewhat different (Mitleb, 1981; Port & O'Dell, 1986; Port & Crawford, 1989). In fact, essentially this same observation had been independently made in other labs for over 20 years (see above papers for references). Of course, negative results are always easy to obtain: e.g., perhaps not all speakers will exhibit difference in the parameters chosen for the experiment, and since the effect is on noisy variables (like vowel and consonant duration), one must clearly have plenty of tokens to study. But statistically significant differences have been demonstrated for production data many times. The perception result is admittedly less well verified. But given the well-attested production differences, it would not seem to be an implausible finding.

So production data show small but statistically significant differences in many measurable properties between the two underlying stops. The differences are typically in the direction one might expect from examination of obviously contrastive pairs like *Bunde–bunte.* Thus the [bʊnt] that is related to underlying /t/ has a slightly shorter preceding vowel and slightly longer closure than the [bʊnt] related

to lexical /d/. Also, there is weaker pulsing during the stop closure and a somewhat stronger release burst on the stops—both of which imply that speakers tend to pull their vocal folds somewhat wider apart for the underlying voiceless stops than for the voiced ones. Of course much overlap remains between the two voicing sets for any of these measurements. But the word pair populations are clearly only slightly different, suggesting that the ''underlying'' voicing feature is still biasing the phonetic detail of the stops despite the fact that most of the difference between the voicing classes has been wiped away when the stop or fricative occurs in this syllablic position.

But can listeners make any perceptual use of these differences? We found in the 1989 paper that if one records a number of productions of this pair and asks native German listeners to identify them (in a forced choice between, e.g., *bunt* and *Bund*), you find that—rather than 50% correct identification (indicating mere guessing due to the supposed neutralization)—roughly 60–80% correct identification of the word pair is obtained (depending, of course, on many linguistic and channel variables that affect the rate of correct identifications). But if subjects were instead identifying a typical minimal pair, such as *Bunde vs. bunte,* in these experimental conditions, they would probably identify them around 99% correct. So the first thing that is puzzling about these data is that the traditional view about phonetics predicts performance close to either chance (=50% in this two-alternative case) or close to 100%. After all, two sounds are either exactly the same or else they must differ by at least one ''phonetic quantum''. And one should expect that any difference will be at least reliably detectible. If such a difference were not reliably detectible, then how could children learn their native language? And how could linguists do phonetic transcriptions that capture ''all and only'' the linguistically controllable phonetic differences?

Adding further strangeness to this case of moderate performance is, as AMR notes, that many well-known phoneticians and phonologists (who should be expected to be at least as good at phonetic discrimination as ordinary lay speaker/hearers) have asserted that the German voicing contrast is ''neutralized'' to some kind of stop that is the same for both *Bund* and *bunt.* Not only did the great German phonetician, E. Sievers, consider the neutralization to be a complete neutralization, but so did L. Bloomfield (1933), W. Moulton (1962) and many others. The D simply turns into a T when it occurs in final position in a syllable. So AMR is understandably puzzled at the paradox: ''How could this difference be fairly easy for native speakers but seemingly be completely impossible for (or perhaps negligently overlooked by) highly respected linguists and phoneticians?'' And the problem is compounded by claims of similar phenomena in syllable-final devoicing in Russian, Polish, Catalan, etc.

In fact, for me personally, one reason why the whole phenomenon seemed quite plausible is that soon after serendipitously discovering this phenomenon with Fares Mitleb, working on his 1981 dissertation, I noticed a very similar effect within my own speech! For me, as in most dialects of American and British English, a similar partial neutralization is found for the flapped medial /t/ and /d/ at least after certain vowels. Thus, for me, pairs like *budding–butting, biddy–bitty* and *ladder–latter* all seem to sound the same (although for some Americans the last pair may exhibit a more noticeable difference in the pronunciation of the [æ] before /d/). Indeed in teaching introductory phonetics classes I had taught students for years that these

pairs were neutralized! ''Say them'', I would say, ''You can hear that they sound the same despite the spelling difference. This is neutralization'', and the students would nod in agreement. So although *bud* and *butt* are obviously contrastive, *budding* and *butting* seem at first to be indistinguishable.

However, Fox & Terbeek (1977) showed that some Americans produce these two ''flaps'' with somewhat different timing (the preceding vowel tends to be shorter before the underlying /t/). In unpublished results, my students and I have found that in some dialects (e.g., New York City), the contrast appears to be completely neutralized except after certain vowels (like [aj]) while in various other dialects differences may be greater and more widespread across the vowels (Port, 1976; Huff, 1980; Chin, 1986). With isolated words read aloud from a list, I find it fairly easy to demonstrate in the classroom that American listeners can guess with much better than chance accuracy whether *budding* or *butting* was intended (using myself as speaker).

So the English flapping situation and German syllable-final devoicing are rather similar. The primary difference is that in English both the /t/ and /d/ are modified (in the pre-unstressed, intervocalic context) to a third sound, the apical flap, whereas in German the voiced obstruents seem to merge directly into the voiceless ones. Both cases are effective enough as neutralizations that native speakers (like phoneticians) do not immediately notice the difference. And in both cases, there is morphological support for the underlying contrast since pairs like *Bund–bunt* and *budding–butting* are only homophonous in certain morphological contexts. Again, in both languages there exist cases where there is no basis beyond orthography for choosing the underlying consonant (cf. German *und* and English *water*).

In the next section I take a closer look at the traditional view of phonetics as presented by Chomsky & Halle (1968) in order to show why this raises serious theoretical issues.

3. Linguistic phonetics: the standard theory

AMR takes the traditional linguistic approach to speech perception very seriously. This theory of linguistic phonetics is essentially the one presented in Chomsky & Halle's 1968 *Sound pattern of English* (SPE, especially pp. 293–301) and is derived from earlier feature theories (Jakobson, Fant & Halle, 1952; Hockett, 1955). Although many specifics of phonology have changed over the years, it is difficult to see much change in the treatment of phonetics from within generative phonology since 1968. Only recently in ''gestural phonology'' (Browman & Goldstein, 1986, 1995) and in the movement toward ''laboratory phonology'' has fundamental change occurred. I will try to summarize this standard theory of phonetics as viewed by many linguists.

3.1. *Competence* vs. *performance*

The whole issue derives from the Chomskyan distinction between *Competence* and *Performance.* When it comes down to mechanisms, this is normally interpreted to differentiate between Competence as the domain of *discrete variables* (that is,

symbols and symbol structure) as they are reconfigured and processed in *discrete time.* Processing time involves discrete jumps between system states when a rule is executed. (The structure of events in real time associated with the pronunciation of words is also discrete but is encoded as the ordering of static objects like segments, words and other syntactic units.) In opposition to this is Performance, the domain of *continuous variables* evolving in *real* (*continuous*) *time*. The continuous variables include processes related to motor control, audition and speech perception. So Performance is essentially the physical, while Competence is cognitive or mental and is assumed to work on principles similar to those of logic, mathematical proof, and digital computation (see Haugeland, 1985; van Gelder & Port, 1995). Within linguistics, it is an article of faith that language (and probably everything else that is mental) will be best understood in terms of discrete time and discrete symbols.

But the competence-performance distinction (closely related to the Mind *vs.* Body distinction) creates problems at both output and input. The first is how can discrete, static symbols independent of real time (in the mind) control real-time continuous performance (in the body)? Presumably the mind controls the body that produces the speech gestures. But, as pointed out by Fowler, Rubin, Remez & Turvey (1981) and Turvey (1990), any model of this process must be quite implausible. The problem is that temporal specifications must now be set for every one of those timeless symbols at output time and then performed somehow. It is not so difficult to postulate rules to specify the durations (so called ''temporal implementation rules'') but it is very difficult to imagine how these specifications could actually be performed in a manner which is not arbitrary. In trying to do it, one is forced to continue discovering new static states (since there is almost no end of subtle contextual effects that can be found), thus causing the problem to blow up exponentially (see, e.g., Port, Cummins & McAuley, 1995). Then some executive system from within performance must assume responsibility to assure that each minisegment type actually lasts the specified amount of time. The second problem, the input problem, is how can the auditory system become a phoneticizer and translate continuous-time auditory events into discrete static symbols? Mechanisms capable of this can be easily constructed, but how could they be designed so as to exploit all the many kinds of subtle temporal information that human speakers and hearers employ? Here too, new intermediate and context-sensitive states tend to proliferate as soon as one looks closely at any data (see, e.g., Port & Rotunno, 1979).

3.2. *Phonetic theory as alphabet*

Of course, Chomsky & Halle did not need to address these performance problems in order to do phonology. They only needed to have some description of speech articulation and speech perception that was sufficient to characterize the sound systems of human languages. They assumed, naturally, that this would take the form of a list of symbols, an inventory. So they asserted the existence of an *interface alphabet,* the list of ''the phonetic capabilities of man'' as they called it in their ringing but now quaintly old-fashioned turn of phrase. These minimal phonetic objects are atomic symbols as far as Competence is concerned, even though within Performance it is assumed they have both articulatory and auditory aspects

involving continuous variables and continuous time—very difficult problems that were left to the phoneticians to deal with.

Thus, on the motor side, these discrete phonetic objects can be thought of as providing a universal alphabet of control configurations—all that is necessary for speech production specification and all that could in principle be controlled by the grammar of a language. Nothing beyond these units (that is, no further articulatory or acoustic detail) is supposed to be controllable—at least, not by the grammar of a language (though apparently people can mimic each other in nonlinguistic ways). On the perception side, it is assumed that audition comes with a ''phoneticizer'' that exhibits ''categorical perception'' and translates continuous acoustic events into discrete symbolic descriptions. It is these two devices, the input device and the output device, that are responsible for phonetic discreteness. And it is their performance that is thrown into question by incomplete neutralization (as well as by many other data, of course).

The traditional linguistic theory of phonetics and speech perception can be summarized this way:

1. Humans can hear the sounds of human speech (only) in terms of a set of discrete sound categories usually called ''phones'';
2. Each phone is a simultaneous combination of phonetic features;
3. Each phonetic feature is a static, atomic object with a simple, unitary articulatory and auditory specification and no internal temporal structure;
4. There is some closed set of such features for human language;
5. Individual languages employ subsets of the universal set for their lexico-phonological systems,
6. Children learning their language employ these units to organize the perception of speech and the phonological grammar of the language.

It follows that two phones may be either identical or distinct. If they have the same phonetic features, then they should be phonetically identical—that is, as far as linguistic control is concerned. (Wouldn't this imply that no human should be able to distinguish them perceptually?) If they are different (that is, have distinct phonetic features), then, at the very least, language learning children and native speakers should be able to differentiate them easily and reliably. Why? Because if children could *not* be counted on to make the appropriate distinctions for any language, then how could various languages be reliably and accurately acquired? This criterion prevents the theoretical linguist from simply enlarging the phonetic alphabet without limit. So on close examination, one discovers that according to the standard theory, the stability of the universal inventory of phonetic units is what provides the explanation for the universal stability of language acquisition.

3.3. *Acquisition*: *adults* vs. *children*

It is well known that many sound units in a non-native language may be quite difficult to acquire when learned by adults. So one must suppose that adults may fail to distinguish the novel sounds of languages due to having somehow lost much of their innate ability to recognize the sound distinctions of human speech (Werker & Tees, 1984; Lively, Pisoni, Yamada, Tohkura & Yamada, 1994). Apparently for

most lay speaker-hearers, only the phonetic categories used in their native language are normally usable for speech perception as adults.

So what about phoneticians and linguists? How do they evade this phonetic atrophy? Presumably the loss of general phonetic resolution may be reduced by phonetic training. The International Phonetic Association (IPA) alphabet and Chapter 7 of SPE are two examples of scientific attempts to organize and list the full set of controllable aspects of speech perception and production. These are supposed to be all that is potentially under linguistic control—whether for contrasting words or for simply controlling the motor system. Linguists and phoneticians cultivate the distinctiveness of these features so as to produce and perceive them. It would seem that a strict version of the standard phonetic theory should predict that humans could never hear *more* detail than what is provided by the "phoneticizer" (though little is normally made of this). An implication of the notion of a universal phonetic alphabet is that we may hope that professional linguists and phoneticians should be able to *approach* the sum of perceptual and motor skills of native speakers of all languages.

3.4. *Formal theories in linguistics*

The Chomsky–Halle model recognized that beyond Competence there is Performance, and acknowledged that the physical signal supporting speech perception is characterized by continuous change over time (e.g., formant trajectories that result from an articulatory motion) and by continuously variable parameters (such as formant frequencies, intensities, lip positions, etc.). However, Chomsky & Halle (C & H) are very clear that *the only aspects of continuous speech events that could be relevant to linguistic competence are those differences that reflect distinct phonetic transcriptions.* The grammars of specific languages can only use some specific universal list of phonetic elements.

The fundamental reason for making this bold assumption is the one pointed out by Haugeland (1985, pp. 52–58): symbolic theories simply must assume a set of positively identifiable symbols. That is, the formal system itself—the grammar—must have symbolic objects that are discrete. They must be discrete in order to be infallibly recognizable. The symbols must also be stable over indefinitely long periods of time: if you put a symbol somewhere in memory, it must still be there when the system comes back later to read it. Formal models depend on these properties in order for their rules to function at all and for data structures to literally hold themselves together. In the execution of a computer program (one familiar example of a formal system), these stabilities are assured due to the engineering of the chip. For human cognition, if it is to be a genuine "competence model" as C & H clearly intend it to be, then these properties must be assumed. Without discreteness, infallible recognition and indefinite time stability, computational models simply will not work. Rules cannot be executed if the system cannot be sure when it is looking at an A rather than a B. In short, *formal linguistics as we know it cannot be done without the assumption of discrete phonetic symbols.*

So C & H had to propose that there is some phoneticizer that chops messy speech into usable symbols. It is assumed that only the output of the speech perception mechanism is available to any language, and this output must be constrained to provide only atomic and static phonetic features selected from the universal set.

This theory, the standard theory of linguistic phonetics, has remained essentially unchanged in the phonological literature since the mid-1960s although it is probably fair to say that C & H formalized ideas that were current from the 1930s on. The most fundamental problem with competence is that within competence there is a sharp distinction between *Symbols* (as indefinitely time-stable states of the system) and the instantaneous *Transitions* between states. This idealization of processing within competence acts as though time is nonexistent! Real time exists neither in the Symbol (which is static) nor in the Transition (which is instantaneous). But real time moves inexorably and can always be looked at over much shorter (or much longer) time scales. Events that appear ''instantaneous'' to our cognitive intuitions may look very slow at the much faster time scale of neurons (just as global neuron behavior looks slow relative to even faster processes like ion channel activity). The Competence-Performance distinction thus amounts to an assumption that whatever might be happening at any shorter time scale within Performance cannot be relevant in any way for what happens at the longer Competence time scale. This is a bold yet almost unexamined assumption (see van Gelder & Port, 1995; Kelso, 1995).

By idealizing Symbols and Transitions in this way, computational models of language make continued scientific progress on a theory of language very difficult. A theory of phonology (and of linguistics as a whole) that can be incorporated into modern cognitive science must begin with a far more sophisticated view of the relation between cognition and the physical aspects of the body and the physical world than merely the simple mapping of performance symbols onto competence symbols offered by an interface alphabet. Insisting on a mere mapping relation between Competence and Performance makes it impossible to understand how language is situated in a nervous system. A practical approach should not assume that linguistics is the study of the symbolic and formal structures of languages, but rather it should view linguistic structures of all kinds, from phones to words to sentences, as events in time. Different kinds of structures ''live'' on different time scales (e.g., sentences are longer than phonemes). Of course, in modern times we have the technology to write words down on paper or put them in a computer file. We can even sample sound waves and put them in a file as well. Then we scan these displays in both directions looking for patterns. But such a display can not be assumed to be available—at least not *a priori*—to human cognition. If such a spatial display of words and sentences does exist cognitively, then accounting for how it could work is an empirical problem. However, to assume that such representations exist seems theoretically reckless, since there is no direct evidence for it whatever (Port *et al.,* 1995). As used by human speakers and as experienced by cognitive systems, *the true dimensional axis of language is time, not space.*

It is the questionable assumption that ''language is a formal symbolic system'' that forces phonology to insist that linguistic phonetics must provide discrete universal objects: phonology needs something formal to manipulate. Of course, there are also a number of other arguments for discreteness that have appeared along the way. Although they are often put forth as relevant evidence by both phonologists and phoneticians, none of these empirical arguments, in my opinion, has more than tangential relevance to the central issue. Nevertheless, it is worth considering these performance-related arguments that seem to be related to the assumption of discrete phonemes.

4. Empirical arguments for phonetic discreteness

There are three main arguments that seem to be proposed most often for why we should believe that the phonetic discreteness assumption and the identifiability assumption are justified. They are that (a) the *limited perceptual resolution* of humans forces a limit on the number of distinctions possible along any dimension and thus supports discrete categories, (b) the phenomenon of *categorical perception* is well attested and suggests that discrete categories with sharp perceptual boundaries are intrinsic to speech perception, and (c) evidence for the *quantal theory of speech* further shows that discreteness is natural for many phonetic categories. I think that none of these provides any compelling evidence that the phonetic space is a universal and invariable discrete inventory.

4.1. *Limited perceptual resolution argument*

To many phonologists the discreteness of speech perception seems almost to require no defense beyond common sense. After all, so the reasoning goes, one can only distinguish so many differences with a certain level of detail. So it seems that only a certain number of, say, vowel distinctions should be possible due to finite limits on resolution on auditory sensation. But is this reasoning sound? Limits on resolution do not necessarily yield discrete levels. In the late 19th century similar reasoning led early psychologists like Titchener and Wundt to similar conclusions about simple stimulus scales such as pitch, color, and so forth (Boring, 1942).

For example, Titchener and others viewed pitch perception as reflecting a sound unit called the ''Just noticeable difference'' (JND). The idea was that the frequency scale for pure tones, for example, must be divided into discrete steps (just like pixels on a computer screen but in one dimension). Thus if two tones are presented to a subject serially and are close enough along the sensory scale to fall within the same JND, then it was predicted that they should be reported as the same, but if they lie in different JND regions, then they should be reported as different. This is quite similar to the kind of reasoning that Chomsky & Halle employed in concluding that vowels must have a fixed (small) number of discrete height values, and similar to AMR when he supposes that either *Bund* and *bunt* are perceptually the same or they are different.

But psychology abandoned the JND view of pitch resolution long ago. The primary reason is the ubiquity of noise internal to the perceptual system. Thus it is clear that listeners, even for a very simple discrimination task, do not always give the same response when the discrimination is difficult. So if you ask them 10 times about the same pair of stimuli, you will often get some ''sames'' and some ''differents''. If one begins with an impossibly small difference and increases the stimulus difference toward easier discriminations and plots the probability of saying ''different'' (from 0 to 1) against the stimulus continuum, the data will always sketch out an S-shaped curve—with higher probability of saying ''different'' corresponding, of course, with larger changes in the stimulus. In fact, if you don't get a smooth curve, then you have either not been sampling closely enough along the stimulus continuum or else have not looked at enough tokens (either within or between listeners).

Can one locate the boundaries between the hypothesized discrete sensory categories? For a small enough difference, moving the difference along the

continuum, one should (on the discrete category view) find discrimination flat spots (where the sounds are within the same sensory class) alternating with discrimination bumps (across a boundary between the sensory steps). But they are not found. These days, when a psychophysicist speaks of a ''just noticeable difference'', it is interpreted to mean enough difference that subjects have, say, a 75% chance of detecting the difference. The S-shaped psychometric function rules out any nonarbitrary steps along stimulus continua. Of course, this is just as true of speech stimuli as for anything else. So the fact that there are sensory limits relevant to distinguishing phonetic categories from each other in no way justifies a claim that there is a discrete set that can be reliably identified.

The notion of ''reliable identification'' runs into another difficulty as soon as probabilistic judgment appears. It turns out that one must differentiate the *sensory analysis* aspect of the discrimination task (or the identification task) from the *response decision* aspect of the problem. Subjects may have a bias toward one response over the other. For example, if subjects are asked to make a discrimination that is sufficiently difficult that subjects won't always give the same response, then other criteria will play a role in determining which response they choose. These are usually called ''response biases''. For example, if the payoffs and penalties of the situation are such that making a ''False Alarm'' (calling them ''same'' when they are not) costs more than the reward for a ''Hit'' (calling them ''same'' when they really are), then observers will tend to be conservative about responding ''same''. Decisions will be affected by a number of features, including the subject's estimate of the *a priori* probability of one state of affairs *vs.* the other (thus, for example, if subjects expect to see more /t/s than /d/s, they will adjust their response criterion to make sure they respond /t/ more often). So, the actual response of subjects (and therefore their actual percent correct in a discrimination or identification task) depends *only in part* on the results of their perceptual analysis of the physical stimulus.

This problem is actually fairly easily solved experimentally: the theory of signal detection (Swets, 1961; tutorial introduction in Kantowicz & Sorkin, 1983) has demonstrated statistical methods to correct for response bias (by taking into account the proportion of hits to false alarms and assuming Gaussian noise distribution) and suggests experimental procedures that permit bias-free measurement of the distinctiveness of two classes.

All of these same features apply to speech. (See Port & O'Dell, 1986 or Port and Crawford, 1989 for simple applications of signal detection theory to the German voicing contrast results.) The best general assumption about the perceptual mechanism is that it produces a probability judgment about a stimulus with respect to several possible categories. Which response a listener (including even a phonetician or linguist) actually gives may reflect a variety of factors that have nothing to do with their perceptual similarity to each category.

4.2. *The categorical perception argument*

It has been known since the 1950s that it you vary speech stimuli along complex continua between phonetic classes, subjects' perception will jump rather discretely (for reviews see Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Repp, 1984; Harnad, 1987). Is this good evidence of a discrete ''phoneticizer'' in speech

perception? No, it is not, for one simple reason. The standard theory of linguistic phonetics requires that *all* phonetic contrasts be discrete while the categorical perception effect has been known from the earliest days to occur more strongly for certain subclasses of sound contrasts (e.g., place of articulation and voicing) than for others, like vowels. In the case of vowels, categorical perception is only obtained under special conditions. Yet discrete categorization of vowels is every bit as critical for vowels as for consonantal features.

Of course, categorical perception is a much more complex problem than distinguishing pure tones differing in frequency. Speech stimuli have enormous complexity and richness, but on the other hand, they receive a huge amount of practice, too. It is pretty clear that when we present listeners with very complex stimuli, only certain aspects of the stimulus tend to be heard accurately (Watson, 1987). Most details cannot be noticed. On the other hand, it is known that if you give subjects a great deal of practice on just a single speech stimulus, listeners can respond in ways that reveal that their auditory resolution approaches the sensory limits observed for simple tones (Kewley-Port, Watson & Foyle, 1988).

So categorical perception is still problematic and not understood, but it is clear that it does *not* provide much justification for assuming that all speech sounds are discretely and reliably perceivable.

4.3. *The quantal theory of speech production argument*

Stevens (1972, 1989) demonstrated that the acoustic response of the human vocal tract behaves highly nonlinearly for certain changes in articulation. The consequence of these nonlinearities is that for speech sounds at certain locations along articulatory continua, any variation in articulatory accuracy will have minimal consequences on acoustics. The ''quantal properties'' of speech suggest that certain places of articulation, certain manners of articulation and certain vowels are relatively insensitive acoustically to articulatory variability. Stevens argued that somehow this justifies or rationalizes the postulation of discrete phonetics. But this evidence really only supports the claim that, given some reasonable assumptions about articulatory and auditory preferences, certain speech sounds may be more ''attractive'' than others. That is, because of these nonlinearities certain sounds may be more efficient choices for languages to employ than others. It explains why certain particular sounds, like [s] and [d] and [a], appear in language after language since these sounds may be both articulatorily and auditorily advantageous. However, it doesn't even begin to provide empirical support for the claim of the standard theory that there is a discrete, reliably identifiable sound inventory innately embedded in human cognition.

In short, none of the empirical, performance-based arguments is directly relevant to the claim that there is a universal, reliably identifiable phonetic alphabet. The fundamental rationale for such an alphabet is really only the original theoretical motivation—that the study of competence phonology cannot even begin without such an alphabet.

Many phonologists would like to believe that experimental research justifies this assumption, but it does not. In fact, data from a century of phonetics research shows that human speech perception is unreliable and nondiscrete. And speech production is the same. It contains enough noise that speakers' productions of the same

linguistic units always span some range if careful measurements are made, and speakers clearly have control over continuous variables that permit sounds to exhibit distributions that overlap to any degree—from statistical identity to being obviously very different. There is no reason to believe that all these difficulties are swept away by some ''phoneticizer''.

One might suppose that even if all this is true, there is still no reason to suppose that anything important has been lost by the assumption of segmental transcription as the basis for phonology. Why can't phonology proceed just fine without any assumptions about discrete phonetics? Whatever the theoretical niceties, one might hope that perhaps there are only rare practical consequences.

5. Consequences for phonology: what is lost?

I am suggesting a rather sceptical view of the process of data collection that is most often employed in phonology. It seems that this process is roughly that linguists (and sometimes phonetic specialists) produce phonetic transcriptions that supposedly represent in symbolic form all the ''linguistically relevant aspects'' of speech production and perception. Then phonologists employ these transcriptions as their primary data for generating models of language-specific data structures of the language. Although these days there are many 'laboratory phonologists' who combine careful data collection on speech sound and on motor control (e.g., Kingston & Beckman, 1990), there remain many phonologists who feel that their work addresses phonological questions for which impressionistic phonetic transcriptions will serve as completely appropriate data. Perhaps. But in my view all phonologists should be at least a little concerned about this assumption.

After all, incomplete neutralization is just one of a vast variety of phenomena on human speech observed experimentally since the second world war. These results show, in broad outline, four conclusions, that:

- There are many kinds of *subtle context effects* that appear in the study of speech articulation or speech acoustics (to get a smattering, see, e.g., Cole, Rudnick, Zue & Reddy, 1980, on spectrogram reading; Liberman *et al.,* 1967; Labov, 1972; Port & Rotunno, 1979);
- Many of these context effects are *manifested in the time domain* (e.g., Lisker & Abramson, 1971; Lehiste, 1970; Klatt, 1976; Port, 1981; Port, Dalby & O'Dell, 1987); and
- Most of these effects are *specific to particular languages*, and cannot be phonetic 'universals' (e.g., Port, Al-Ani & Maeda, 1980; Port, Dalby & O'Dell, 1987).

It is these subtleties that have made automatic speech recognition so challenging. Furthermore;

- Essentially any of these subtle variables, under appropriate conditions, can be *employed by listeners in speech perception* (too many references to cite, but intriguing cases include Dorman, Raphael & Liberman, 1979; Port, Reilly & Maki, 1988; Port, Mora & de Jonge, 1992).

Thus it is very likely that there are a great many aspects of languages that are phonologically important (since they differ from language to language) yet are

completely missed due to reliance on the traditional symbolic phonetic transcription of speech.

Let me mention briefly here two specific examples of phonological phenomena in areas familiar to me that slip through ''the segmental grid''—problems that seem to be ignored or dealt with awkwardly in phonology primarily due to reliance on discrete phonetic transcriptions.

5.1. *The Germanic postvocalic* [*voice*] *contrast*

In English when the voicing contrast occurs at the end of a syllable, as in the pair *fuzz* and *fuss,* the difference in voicing is manifested as a change in the ratio of the duration of the vocalic part of the syllable to the duration of the final consonantal portion of the syllable (Port, 1981; Port & Dalby, 1981). This durational ratio also helps to characterize the contrast between, say, *bids–bits, camber–camper, Libby–lippy, Bangor–banker, large–larch, ruby–rupee,* etc. A similar contrast in the ''vowel/consonant duration ratio'' for distinguishing voicing pairs is also found in many other Germanic languages (at least Standard German, Bavarian, Swedish, and Icelandic). But if you do just a segmental transcription to represent the data, then the durational difference between the stop and fricative closures (e.g., between [t] and [d] or [s] and [z]) seems uninteresting (because it is said to affect only ''phonetic implementation'', not the phonology) and the effect on the preceding vowel and any voiced consonant (e.g., the nasal in *lunge–lunch*) is just another instance of a context-dependent phonological rule, of which there are many well-known examples. The compensatory or inverse durational relationship is completely obscured—due entirely to the restriction to segmental transcription.

So here is an important language-specific phenomenon, with many variants across the Germanic family (including the very ''incomplete neutralization'' phenomenon that stimulated AMR's essay). This shortening followed by lengthening (taking [−voice] to be derived from [+voice]) could be viewed as a brief perturbation of speaking rate (see Port & Cummins, 1992, for such an interpretation). But however this ratio effect should be described, it is clearly phonology since it is part of the grammar of English, German, Icelandic, etc. Most other languages do not show evidence of manipulation of these temporal ratios as a correlate of a voicing-like contrast. This seems to be, on the face of it, a fascinating phonological problem. But it lies in the time domain and is generally overlooked.

5.2. *English* ''*meter*''

A second domain where the description of speech in symbolic terms may render important phonological structures invisible is in the problem of meter. English phrases often seem to have a global timing structure over a scale of a second or so. Phrases like *Mississippi legislators* seem to have an alternating pattern of stresses—whether four of them or only two (Hayes, 1995). It has been proposed many times (Jones, 1932; Abercrombie, 1967; Martin, 1972) that music-like rhythm, definable in terms of relative duration, may underlie such pronunciations. Unfortunately, actual temporal studies typically find messy and unclear results (see Lehiste, 1977; Port, Cummins & Gasser, 1996). Consequently, phonologists will often address the problem of meter with a discrete time scale, using one time step for every syllable (e.g., Halle & Vergnaud, 1980; Hayes, 1995). But will discrete

time prove sufficient for an understanding of English metrics? Certainly it cannot provide a complete understanding, since production and perception always take place in real time.

In recent experiments in my lab, we have been exploring the temporal aspects of English metrics with some new methods (Cummins & Port, 1996*a*, 1996*b*; Port, Cummins & Gasser, 1996). We first hypothesized that some sort of real-time oscillatory system might underlie the metrical aspects of speech timing. If this is so, then we should be able to interfere with such an oscillatory system by encouraging "coupling" with another oscillatory pattern (inspired by the work of Kelso and by Treffner & Turvey, 1993). To illustrate what is meant by coupling, imagine a parent pushing their child on a swing. The parent will couple their body motions to the rate and amplitude of the oscillating child-swing system. If the swing length changed or if extra weights were put on the seat, then the parent would adapt to the change in frequency—that is, they will remain coupled to the child-swing oscillator. This state of coupling provides the most efficient way for them to use their body to keep the swing going. Coupling is found both within our bodies (e.g., between your left and right legs when walking) and between our own body and that of others (e.g., in communal singing or marching).

Oscillators that are coupled tend to impose very severe constraints on each other's frequency and phase. For example, imagine tapping your finger on the table in a comfortable position. If you are asked to tap the index finger on your left hand at some rate, you could do so at any rate over a broad range from fast to slow. But if you are asked to also oscillate your right index finger at some rate, then it turns out that you will be able to perform both tasks together only at a small set of rates. In fact, they will be rates such that there is a very simple ratio, such as 1:1, 1:2 or 1:3 (or, with some practice, 2:3, 3:4, etc.), between the two fingers. Apparently, two fingers in the same body cannot avoid coupling with each other. They "want" to keep certain simple temporal and phase relationships.

We reasoned that evidence of coupling between a periodic stimulus and human behavior can be interpreted as evidence that the behavioral system incorporates an oscillator. Could we show that the relationship between a metrical foot and the phrase resembles coupled oscillators? In our experiments we asked speakers to listen to a metronome signal (at a comfortable level) and repeat a simple phrase. Thus they might say "*Talk to the boy*" once for each beep of a metronome (at periods from 0.3 s up to 1 s). This phrase has two metrical feet: "*talk to the*" and "*boy*". Not only did we find that speakers align "*talk*" with the metronome pulse (just as we instructed them to do), but the onset of "*boy*" has a very strong tendency to fall at certain phase angles rather than others, especially at 1/2 (but also at 1/3 or 2/3) of the cycle from "*talk*" to "*talk*". (The reader is encouraged to try simply repeating this phrase. You will probably find that the perceptual beat of "*boy*" is located half way between the phrase onsets.) So by means of this simple task—repeating a phrase to a metronome—we demonstrated a strong tendency for the "foot oscillator" to entrain itself with the "phrase oscillator" in an integer ratio like 2:1 or 3:1. We take the ease with which speakers couple their speech to a metronome and the tendency for feet to couple with the longer phrase unit to suggest that "hierarchically nested oscillators" running in continuous time underlie the metrical structure of, at least, English—whether or not there happens to be a metronome to couple with. Otherwise this coupling with the metronome should

not be so easily obtained. If there is another periodic action—e.g., if you are also tapping your finger, pounding your fist, marching, jogging, talking, chewing gum, or whatever—then your speech will tend to couple with it. And there is no way for a metrical phonology based on symbol sequences to explain coupling with real-time periodic events since symbol sequences involve no real time at all.

Of course, none of this implies that traditional metrical phonology, such as the work of Hayes (1995), using discrete time as the basis for meter, is not worthwhile. However, it seems that an empirically adequate understanding of meter will come only when the insights from discrete-time descriptions can be understood or reinterpreted in terms of a dynamical model of meter for English. When that is attempted, I suspect that some current issues will lose their interest (e.g., ''rhythm-rule'' phenomena will be much more clearly understood) while other phenomena will find insightful new interpretations from the dynamical perspective.

I have argued that reliably identifiable and discrete phonetics is an unavoidable assumption for a formal or competence model of phonology and linguistics, but that no such reliable state-based speech perception is possible no matter how many years of phonetic training you have. But there is still one essential task left to do in this essay. This is to offer a specific account of how the phoneticians could miss something that the native speakers could hear fairly easily.

6. Can phoneticians fail where native speakers succeed?

How could it be that phoneticians differ so sharply from lay native speakers in their ability to perceive this distinction in German? Given all that has been described so far about probabilistic perceptual outputs and about how response choices are affected both by stimulus analysis as well as other criteria, the answer is quite simple. They performed completely different tasks. Phoneticians and linguists are professionally concerned with establishing and applying consistently a set of phonetic units that are plausible candidates for ''the universal phonetic space''. They would be understandably reluctant to claim to have discovered some categorical difference that they can only differentiate with, say, 70% accuracy. That is, given the standard theory of phonetics, there are many criteria aside from the mere perceptibility of a difference that are relevant to their decisions about what phonetic transcription to use. We might say that they have a professionally motivated response bias. Only differences that are large enough and reliable enough are of interest. Thus it is important for them to ignore mere token-to-token variation, speaker idiosyncrasies, minor dialect differences, and so forth. Very likely these phoneticians and phonologists agree with Chomsky & Halle that any distinction that is to be represented in the universal phonetic alphabet must be large enough that it might be useful (in some imaginable situation) within a language to distinguish words with fair reliability. From the perspective of these criteria, there is no question that the two [t]s in *Bund* and *bunt* should *not* be differentiated. They are indeed too similar to be of contrastive use.

On the other hand, in our discrimination experiments with native speaking Germans, the question we asked them was ''Which word did the speaker say?''—rather than ''Which phonetic symbols would you use to describe these sounds?'' So these listeners used all the acoustic evidence they could find to make their guess. There is no question in my mind that any serious phonetician, whether

or not a native speaker of German, could do nearly as well as our German listeners at the word identity task with very small amount of training with feedback (e.g., 10–20 trials). I am not a fluent speaker of German but was able myself to perform as well at the task as our better German listeners. Like our subjects, I knew that there was an equal number of underlying /d/s and /t/s, so I just listened for *more* D-like *vs less* D-like stimuli. The point is that choosing a phonetic transcription and identifying a word are two vastly different tasks with little in common except the stimuli themselves. It is little wonder that the phoneticians ignored these minute differences in assigning their transcription, but neither is it surprising that the native listeners did use this information in identifying the words.

The view within traditional phonology that phonetic transcriptions can be reliably and discretely assigned is a bit naive, it seems to me. What seems to be happening in the field is that the discipline of phonology is in the process of splitting between those who do careful experimental studies of speech and those who insist that impressionistic phonetics, often obtained largely from secondary sources, is good enough.

7. Toward phonological morphogenesis

Thus far my arguments have been largely negative, to insist that there is no reason to assume discrete categorization in phonetics. The implication is to pull the discreteness rug out from under phonology. So if phonetics cannot explain the discreteness of phonology, how can we account for the distinct phonological objects we linguists so clearly observe there? Languages seem to exhibit discretely different places of articulation, manners, vowel heights, etc., in the units used to spell lexical entries. Where could these come from? But first note that linguistics is not unique in trying to account for the nature of structured things, of identities invariant under transformation. Such issues lie at the heart of biology and other fields. Indeed, the notion of an ''object'' or identity turns out to be just as problematic even for computer science (Smith, 1996).

If one assumes that cognition is a continuous-time dynamical process embedded in neural tissue and an external physical environment, then a new set of theoretical tools become appropriate—the tools of dynamical systems (see Kelso, 1995; Port & van Gelder, 1995). Morphological structures arise in this world on many temporal and spatial scales, from the individual stars and galaxies of astronomy, to species and self-organizing skin patterns in biology, to molecular valences or tornadoes in physics. Simple integer ratios are found even in the resonant frequencies of uniform tubes and strings. None of these spatio-temporal forms can be explained by simple ''analysis into their parts'' and cannot be constructed by ''merely assembling together their structural atoms''—like the way that phonological structures are assembled from phonetic atoms. Instead, they seem to generate themselves over time.

René Thom coined the term *morphogenesis* to describe the process of the creation of a form as a temporally stable pattern, typically exhibiting some specific symmetries (often periodic ones in space or time) from within a flux of dissipating energy. The mechanisms by which such self-organizing dynamics create an ''object'' (or other more complex forms) are beginning to be understood at a mathematical level of abstraction (Haken, 1983; Kelso, Ding & Schöner, 1994; Kelso, 1995). If we

abandon the dogma that ''language is a formal system'', we may see that phonological structures, as structured events in time, are phenomena that are not completely unrelated to other domains of biology, and may be explained in similar ways. Cognition is a system that runs in continuous time following dynamical laws. In some situations, it produces discrete ''object''-like structures in space-time. In the case of speech, for example, the relevant space is an abstract one that can be derived from either articulation or acoustics.

Rather than merely providing a specification of parameters that some (yet unimagined) production and perception implementational systems are supposed to carry out in real time, a dynamical description will show (or at least suggest) just how such phonological objects could be both produced and perceived. Obviously it will be a challenging task to develop a theory of phonology along these lines. But it seems clear that this is where the future of our discipline lies. The theory of language must show at least how abstract linguistic structures that are invariant across a community of speakers as well as between production and perception could be created as stable ''morphs'' of a community of speaker-hearers (Port, 1986).

8. Conclusions

This essay began to answer a straightforward question, one that appeared to be primarily about mere facts. But I wanted to provide a good answer to AMRs worthy question about the serious implications of incomplete neutralization rules. There are two main conclusions to be drawn from this discussion.

First, the lack of complete neutralization of syllable-final voicing in German is a real phenomenon. It has been observed many times. There appear to be several ''sounds'' here produced and perceived with only a moderate degree of distinctiveness. But whether they are the same or different depends on how you ask the question. The American English flapping situation is another similar example. There are others as well, such as the near neutralization of pairs like *prints–prince* in American English (Fourakis & Port, 1986). As for other proposed cases in Polish and Catalan, etc., I am not in a position to make bold claims. Nor am I certain, in a dialect exhibiting an incomplete neutralization, that all speakers will necessarily show the effect. Speaker idiosyncrasy seems a possibility here. After all, who is likely to notice which way a particular speaker implements the ''neutralization''?

Secondly, I agree completely with AMR that this effect is similar to other instances of the ''near-merger'' of a contrast (Labov, 1972) and the gradual reduction of contrasts under ''weakening'' and ''fast speech'', etc. But more sweepingly, it seems that 50 years of experimental phonetics research on speech production, speech motor control, speech perception and descriptive phonetics have shown repeatedly that human speech sounds tend to distribute themselves rather smoothly over a wide range of variables. The incomplete neutralization result is just one particularly vivid and carefully studied example in a great mass of data supporting continuous phonetics over discrete phonetics.

AMR is right to be troubled about incomplete neutralization since it undermines a critical assumption about phonetics that is taken to be ''gospel'' by many working phonologists. It seems to me (as it does for many phoneticians and ''lab phonologists'') that, by trusting segmental transcriptions, phonologists run the risk, quite frankly, of building their work on sand. Such work tends to completely ignore

everything about time except for what can be expressed in terms of the serial order of symbols. And they tend to ignore or represent only clumsily the many graded effects like ''weakening'', speaking rate, and ''articulatory laziness''. By ignoring continuous-time effects, phonologists run the risk of developing theories of the wrong phenomena and of overlooking important language-specific phenomena. Languages do exhibit some discrete ''sound objects'' which cry out for description and explanation: ''distinctive features'', ''phonemes'', ''stress levels'', ''natural classes'' (like obstruents, vowels, nasals, etc.) and much more. But these are all phonological objects—*linguistic objects*—and it is a linguistic theory that must find a way to explain them. Alas, phonetics cannot.

One final remark on the practice of phonetic transcription: Despite my insistence that no alphabet for phonetics can be completely relied upon, I continue to teach the IPA alphabet to linguistics students. It is useful for very many purposes—especially for communication about our work. So it does matter that we have an up-to-date, reasonably standardized alphabet for our papers and journals. But I also teach my students not to trust any alphabetic description of speech, and not to imagine that their or anyone else's transcriptions provide reliable descriptive units that capture all of the phenomena for which we linguists seek understanding.

The author is grateful to Fred Cummins, Stuart Davis, Kenneth de Jong, Michael Gasser and to the editor for comments on versions of this manuscript and to Paul Kienzle for other assistance. This research was supported in part by the Office of Naval Research, N0003-1267.

References

Abercrombie, D. (1967) *Elements of general phonetics.* Chicago: Aldine Pub. Co.
Bloomfield, L. (1933) *Language.* New York: H. Holt and Company.
Boring, E. G. (1942) *Sensation and perception in the history of experimental psychology.* New York, London: D. Appleton-Century Company.
Browman, C. P. & Goldstein, L. (1986) Towards an articulatory phonology, *Phonology Yearbook,* **3,** 219–252.
Browman, C. P. & Goldstein, L. (1995) Dynamics and articulatory phonology. In *Mind as motion*: *explorations in the dynamics of cognition.* (Port, R. F. and van Gelder, T., editors), pp. 175–193; Cambridge, MA: MIT Press.
Chin, S. (1986) Flaps in American English. Unpublished manuscript.
Chomsky, N. & Halle, M. (1968) *The sound pattern of English.* New York: Harper and Row.
Cole, R., Rudnicky, A., Zue, V. & Reddy, R. (1980) Speech as patterns on paper. In *Perception and production of fluent speech* (Cole, R., editor), pp. 3–50. Hillsdale, NJ: L. Erlbaum.
Cummins, F. & Port, R. F. (1996*a*) Rhythmic commonalities between hand gestures and speech. In *Proceedings of the eighteenth meeting of the Cognitive Science Society.* To appear.
Cummins, F. & Port, R. F. (1996*b*) Rhythmic constraints on English stress timing. In *Proceedings of the fourth international conference on spoken language processing.* To appear.
Dorman, M., Raphael, L. & Liberman, A. (1979) Some experiments on the sound of silence in phonetic perception, *Journal of the Acoustical Society of America,* **65**, 1518–1532.
Fourakis, M. & Port, R. (1986) Stop epenthesis in English, *Journal of Phonetics,* **14**, 197–221.
Fowler, C. A., Rubin, P., Remez, R. & Turvey, M. (1981) Implications for speech production of a general theory of action. In *Language production* (Butterworth, B., editor), pp. 373–420. New York: Academic Press.
Fox, R. & Terbeek, D. (1977) Dental flaps, vowel duration and rule ordering in American English, *Journal of Phonetics,* **5**, 27–34.
Haken, H. (1983) *Synergetics, an introduction*: *non-equilibrium phase transitions and self-organization in physics, chemistry, and biology.* Berlin: Springer.
Halle, M. & Vergnaud, J.-R. (1980) Three dimensional phonology, *Journal of Linguistic Research,* **1**(1), 83–105.
Harnad, S. (1987) Category induction and representation. In *Categorical perception* (Harnad, S., editor), pp. 1–20. Cambridge, MA: Cambridge University Press.
Haugeland, J. (1985) *Artificial intelligence*: *the very idea.* Cambridge, MA; Bradford Books MIT Press.

Hayes, B. (1995) *Metrical stress theory*: *principles and case studies.* Chicago: University of Chicago Press.
Hockett, C. (1955) *A manual of phonology.* Baltimore: Waverly Press.
Huff, C. (1980) Voicing and flap neutralization in New York City. In *Research in Phonetics,* **1**, 233–256. Indiana University, Department of Linguistics.
Jakobson, R., Fant, G. & Halle, M. (1952) *Preliminaries to speech analysis*: *the distinctive features and their correlates.* Cambridge, MA: MIT Press.
Jones, D. (1932) *An outline of English phonetics.* 3rd edition. 1st edition published 1918. Cambridge, U.K.: Cambridge University Press.
Kantowicz, B. & Sorkin, R. (1983) *Human factors*: *understanding people-system relationships.* New York: Wiley.
Keating, P. A. (1985) Universal phonetics and the organization of grammars. In *Phonetic linguistics*: *essays in honor of Peter Ladefoged* (Fromkin, V., editor) pp. 115–132. New York: Academic Press.
Kelso, J. A. S., Ding, M. & Schöner, G. (1994) Dynamic pattern formation: a primer. In *A dynamic systems approach to the development of cognition and action* (Smith L. B. and Thelen, E., editors), pp. 13–50. Cambridge, MA: Bradford Books MIT Press.
Kelso, S.(1995) *Dynamic patterns*: *the self-organization of brain and behavior.* Cambridge, MA: MIT Press.
Kewley-Port, D., Watson, C. & Foyle, D. (1988) Auditory temporal acuity in relation to category boundaries: speech and nonspeech stimuli, *Journal of the Acoustical Society of America,* **83**, 1133–1145.
Kingston, J. & Beckman, M. E., editors (1990) *Papers in laboratory phonology I*: *between the grammar and physics of speech,* New York: Cambridge University Press.
Klatt, D. (1976) Linguistic uses of segmental duration in English: acoustic and perceptual evidence, *Journal of the Acoustical Society of America,* **59**, 1208–1221.
Labov, W. (1972) *Sociolinguistic patterns*: *the study of language in its social context.* Philadelphia: University of Pennsylvania Press.
Lehiste, I. (1970) *Suprasegmentals.* Cambridge, MA: MIT Press.
Lehiste, I. (1977) Isochrony reconsidered, *Journal of Phonetics,* **5**, 253–263.
Liberman, A., Cooper, F., Shankweiler, D. & Studdert-Kennedy, M. (1967) Perception of the speech code, *Psychological Review,* **74**, 431–461.
Lindblom, B. (1983) Economy of speech gestures. In *The production of speech* (MacNeilage, P., editor), pp. 217–246. Berlin: Springer.
Lisker, L. & Abramson, A. (1971) Distinctive features and laryngeal control, *Language,* **44**, 767–785.
Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. & Yamada, T. (1994) Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories, *Journal of the Acoustical Society of America,* **96**, 2076–2087.
Manaster Ramer, A. (1996) A letter from an incompletely neutral phonologist, *Journal of Phonetics,* this issue.
Martin, J. G. (1972) Rhythmic (hieararchical) versus serial structure in speech and other behavior, *Psychological Review,* **79**, 487–509.
Mitleb, F. M. (1981) *Segmental and non-segmental structure in phonetics*: *evidence from foreign accent.* PhD thesis, Department of Linguistics, Indiana University, Bloomington, Indiana.
Moulton, W. (1962) *The sounds of English and German.* Chicago: University of Chicago Press.
Port, R. (1976) *The influence of speaking tempo on vowel and consonant duration in English.* Bloomington, IN: Indiana University Linguistics Club.
Port, R. F. (1981) Linguistic timing factors in combination, *Journal of the Acoustic Society of America,* **69**, 262–274.
Port, R. (1986) Invariance in phonetics. In *Invariance and variability in speech processes* (Perkell, J. & Klatt, D., editors), pp. 540–558. Hillsdale, NJ: Erlbaum Associates.
Port, R. F., Al-Ani, S. & Maeda, S. (1980) Temporal compensation and universal phonetics, *Phonetica,* **37**, 235–252.
Port, R. & Crawford, P. (1989) Pragmatic effects on neutralization rules, *Journal of Phonetics,* **16**, 257–282.
Port, R. & Cummins, F. (1992) The English voicing contrast as velocity perturbation. In *Proceedings of the* 1992 *international conference on spoken language processing* (Ohala, J., Nearey, T., Derwing, B., Hodge, M. & Wiebe, G., editors), pp. 1311–1314, Edmonton, Alberta: University of Alberta.
Port, R., Cummins, F. & Gasser, M. (1996) Dynamic approach to rhythm in language: Toward a temporal phonology. In *CLS* 31: *papers from the* 31*st meeting of the Chicago Linguistics Society* (Dainor, A., Hemphill, R., Luka, B., Need, B. & Pargman, S., editors), pp. 375–397, Chicago, IL: Chicago Linguistics Society.
Port, R. F., Cummins, F. & McAuley, J. D. (1995) Naive time, temporal patterns and human audition. In *Mind as motion*: *explorations in the dynamics of cognition* (Port, R. F. & van Gelder, T., editors), pp. 339–371. Cambridge, MA: MIT Press.
Port, R. & Dalby, J. (1982) C/V ratio as a cue for voicing in English, *Journal of the Acoustical Society of America,* **69**, 262–74.

Port, R. F., Dalby, J. & O'Dell, M. (1987) Evidence for mora timing in Japanese, *Journal of the Acoustical Society of America,* **81**, 1574–1585.
Port, R., Mora, J. P. & deJonge, C. (1992) Usefulness of temporal detail for word identification by native and non-native listeners, *Research in Phonetics,* **6**, 147–163. Indiana University, Department of Linguistics.
Port, R. & O'Dell, M. (1986) Neutralization of syllable-final voicing in German, *Journal of Phonetics,* **13**, 455–471.
Port, R. F., Reilly, W. T. & Maki, D. P. (1988) Use of syllable-scale timing to discriminate words, *Journal of the Acoustical Society of America,* **83**, 265–273.
Port, R. F. & Rotunno, R. (1979) Relation between voice-onset time and vowel duration, *Journal of the Acoustical Society of America,* **66**, 654–662.
Port, R. F. & van Gelder, T., editors (1995) *Mind as motion*: *explorations in the dynamics of cognition.* Cambridge, MA: Bradford Books MIT Press.
Repp, B. (1984) Categorical perception: issues, methods and findings. In *Speech and language*: *advances in basic research and practice,* **10***,* (Lass, N. J., editor) pp. 243–335. Hillsdale, NJ: Erlbaum Associates.
Smith, B. C. (1996) *On the origin of objects.* Cambridge, MA: Bradford Books MIT Press.
Stevens, K. N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In *Human communication*: *a unified view* (David, E. E. & Denes, P. B., editors), pp. 51–66. New York: McGraw-Hill.
Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics,* **17**, 3–45.
Swets, J. A. (1961) Is there a sensory threshold? *Science,* **34**, 168–177.
Treffner, P. J. & Turvey, M. T. (1993) Resonance constraints on rhythmic movement, *Journal of Experimental Psychology*: *Human Perception and Performance,* **19**, 1221–1237.
Turvey, M. T. (1990) Coordination, *American Psychologist,* **45**, 938–953.
van Gelder, T. & Port, R. (1995) It's about time: overview of the dynamical approach to cognition. In *Mind as motion*: *explorations in the dynamics of cognition* (Port, R. F. & van Gelder, T., editors) pp. 1–43. Cambridge, MA: Bradford Books MIT Press.
Watson, C. S. (1987) Uncertainty, informational masking, and the capacity of immediate auditory memory. In *Auditory processing of complex sounds* (Yost, W. A. & Watson, C., editors) pp. 267–277. Hillsdale, NJ: Erlbaum Associates.
Werker, J. & Tees, R. (1984) Phonemic and phonetic factors in adult cross-language speech perception, *Journal of the Acoustical Society of America,* **75,** 1866–1978.